AFCAPS-FR-2012-0005

Enlisted Selection and Classification Tests: Precursors of the ASVAB

Nancy Thompson
Operational Technologies Corporation
4100 N. W. Loop 410, Suite 230
San Antonio, TX. 78229-4253

June 2007

Sponsored by:

Kenneth L. Schwartz
Johnny J. Weissmuller
Air Force Personnel Center
Strategic Research and Assessment
HQ AFPC/DSYX
550 C Street West, Ste 45
Randolph AFB TX 78150-4747

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report was cleared for release by HQ AFPC/DSYX Strategic Research and Assessment Branch and is releasable to the Defense Technical Information Center. The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report is published as received with minor grammatical corrections. The views expressed are those of the authors and not necessarily those of the United States Government, the United States Department of Defense, or the United States Air Force. In the interest of expediting publication of impartial statistical analysis of Air Force tests SRAB does not edit nor revise Contractor assessments appropriate to the private sector which do not apply within military context.

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct request for copies of this report to:

Defense Technical Information Center - http://www.dtic.mil/

Available for public release. Distribution Unlimited. Please contact AFPC/DSYX Strategic Research and Assessment with any questions or concerns with the report.

This paper has been reviewed by the Air Force Center for Applied Personnel Studies (AFCAPS) and is approved for publication. AFCAPS members include: Senior Editor Dr. Thomas R. Carretta AFMC 711 HPW/RHCI, Associate Editor Dr. Gregory Manley HQ AFPC/DSYX, Dr. Lisa Hughes AF/A1PF, Dr. Paul DiTullio AF/A1PF, Kenneth Schwartz HQ AFPC/DSYX, Johnny Weissmuller HQ AFPC/DSYX, Dr. Laura Barron HQ AFPC/DSYX, Dr. Mark Rose HQ AFPC/DSYX, and Brian Chasse HQ AFPC/DSYX.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-06-2007	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Enlisted Selection and Classification Tests: Precursors to the ASVAB	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Nancy Thompson	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Operational Technologies Corporation 4100 N. W. Loop 410, Suite 230 San Antonio, TX. 78229-4253.	8. PERFORMING ORGANIZATION REPORT
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) **Air Force Personnel Center Strategic Research and Assessment Branch Randolph AFB TX 78150**	10. SPONSOR/MONITOR'S ACRONYM(S) HQ AFPC/DSYX
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFCAPS-FR-2012-0005

12. DISTRIBUTION / AVAILABILITY STATEMENT
Available for public release. Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes the evolution of selection and classification tests used for enlisted personnel throughout the Air Force's history. It starts with the evolution of aptitude tests used for selection during World War I (Army Alpha and Army Beta), the Army General Classification Test (AGCT) used during World War II, up to the development and use of the Armed Services Vocational Aptitude Battery (ASVAB) that is currently used today. The report also describes the development of the Mechanical, Aptitude, General, and Electronic (MAGE) composites of the Airmen Classification Battery that later turned into the ASVAB. These composites have been used for USAF enlistment classification for the past 50 years.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	U	27	Kenneth L. Schwartz
					19b. TELEPHONE NUMBER *(include area code)* 210-565-3139

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

CONTENTS

TABLES

ENLISTED SELECTION AND CLASSIFICATION TESTS: PRECURSORS OF THE ASVAB

I. INTRODUCTION

The US Air Force uses aptitude tests for selecting recruits who are likely to succeed in their jobs and to help classify individuals into occupational specialties. The better these tests predict performance, the more effectively the Air Force can accomplish its mission and use its resources.

The process of selection is a yes or no decision. Based on selection tests and other factors, an individual is accepted into the Air Force or rejected when the selection factors indicate that the applicant may not be able to perform up to the Air Force standards. Once the applicant is accepted, the issue of assigning the recruit to a job becomes more complex. Every recruit must be placed into a job and given training for that job. The problem is to determine which of many potential military specialties would be the best match for each recruit. For each recruit, the question is whether predicted performance and utility for the Air Force will be optimal when the recruit is assigned as a clerk, mechanic, linguist, or some other occupation. This question takes on a huge magnitude when it must be answered for thousands of recruits in over 100 occupational specialties. Classification tests are designed to measure aptitudes for performance in occupational specialties. Composite scores derived from the combinations of scores from these aptitude tests are used to determine qualifications for these specialties.

This report will discuss the evolution of selection and classification tests for the US Air Force beginning with the first tests developed during World War I through the initiation of the Armed Services Vocational Aptitude Battery (ASVAB) that is used today for enlisted selection and classification. A summary of selection and classification tests for the 50-year period can be found at Table 1.

II. SELECTION AND CLASSIFICATION TESTS OF WORLD WAR I AND WORLD WAR II

The military was a pioneer in the development and use of aptitude tests. As early as 1908, Alfred Binet had suggested using mental testing of conscripts to eliminate those who were considered mental defectives in the Army. The American Psychological Association Committee on Examination of Recruits was instrumental in initiating testing in the military. Robert Yerkes, as a member of this committee, wrote that "we should not work primarily for the exclusion of intellectual defectives, but rather for the classification of men in order that they may be properly placed in the military service" (Wigdor, & Green, Jr., 1991). Directed by the Psychology Committee of the National Research Council, which was established by the National Academy of Sciences in 1916, the committee on examination of recruits developed an intelligence test for screening large

Table 1
Early Air Force Selection and Classification Tests*

Aptitude Test	Date Implemented	Used for Selection	Used for Classification
World War I Army Alpha**	1917		
Army Beta**	1918		
World War II Army General Classification Test (AGCT)**	1940	X	
Post-World War II Airman Classification Battery AC-1A	1948		X
Airman Classification Battery AC-1B	1949		X
Armed Forces Qualification Test (AFQT)	1950	X	
Airman Classification Battery AC-2A	1956		X
Armed Forces Women's Selection Test (AFWST)	1956	X	
Airman Qualifying Examination, Form D (AQE-D)	1958	X	X
Airman Qualifying Examination, Form F (AQE-F)	1960	X	X
Airman Qualifying Examination – 1962 (AQE-62)	1962	X	X
Airman Qualifying Examination – 1964 (AQE-64)	1964	X	X
Airman Qualifying Examination – 1966 (AQE-66)	1966	X	X
Airman Qualifying Examination, Form J (AQE-F)	1971	X	X
Transition to ASVAB Armed Services Vocational Aptitude Battery, Form 3 (ASVAB-3)	1973	X	X

*Includes early World War I and World War II Army aptitude tests. ** Army Alpha, Beta, and AGCT were used for placement decisions. Note: During the years when the AQE was administered, forms of the Airman Classification Battery were used for purposes other than selective enlistment.

groups of men using a multiple choice test format. The resulting test was the Army Alpha which was approved on Christmas Eve 1917 for testing of all enlisted men, draftees, and officers. It was a verbal test administered in a group setting and composed of eight subtests covering verbal, numerical, information, and the ability to follow directions. A second version of the test was developed in 1918 and was called the Army Beta. It was a non-verbal counterpart to the Alpha and designed for use with illiterates and those who could not speak English. The tests were administered to men already in the Army. Based on their scores, recruits were classified as mentally low, mentally average, mentally high, or irregular (Judy, 1969). Test results were used for several purposes including placement of recruits. Many in the military establishment did not approve of the mental testing; but by the end of the war, about 1,750,000 men had taken one of the tests. In 1921, a version of the Army Alpha was published as the National Intelligence Test. The use of tests grew to include different tests for civilian use, including common use in schools across the country (Wigdor, & Green, Jr., 1991). By 1919, the Army mental testing program had been abandoned, but the contribution of mental testing as developed by Yerkes and his team of World War I psychologists continues to have a phenomenal effect in the evaluation of individuals in school, employment, and other areas.

Early in World War II, the standard for induction was "the capacity of reading and writing the English language as prescribed for the fourth grade in grammar school" (Wigdor, & Green, Jr., 1991). During World War II, the military began using aptitude tests for screening and placement purposes. The Army General Classification Test (AGCT) was put into use in 1940 and administered to over nine million Army Air Forces and Marine recruits, and the Navy General Classification Test (NGCT) was administered to over three million sailors (Maier, 1993). Minimum scores on the tests for military qualification fluctuated as the manpower requirements for the war changed. This fluctuation in testing standards was a trend that would continue as manpower needs and military policy changed over the years.

The AGCT became an early standard for aptitude testing. With the start of the war, the Army found that they needed to select individuals who could be trained quickly and eliminate individuals who could not perform in wartime. The AGCT was a test of general learning ability that would help identify those who could perform in wartime situations. It was standardized on a white male military personnel sample and Civilian Conservation Corps enrollees. Scores were given to inductees based on standard scores ranging from the highest of Army Grade I to the lowest of Army Grade V (Maier, 1993).

The AGCT was found to have adverse impact on the Spanish speaking population. It was translated into Spanish in 1941 and was called Examen Calification de Fuerzas Armadas (ECFA). It was later used in place of the Armed Forces Qualification Test (AFQT) in screening for Puerto Ricans. Puerto Ricans continued to receive special testing and language training through the 1980's (Maier, 1993).

The Army Air Forces Aviation Psychology Program also contributed to the development of principles of personnel classification that were applied to enlisted

personnel. This research program focused on developing tests to classify and assign officers to pilot, navigator, and bombardier training. The Army Air Force Qualifying Examination (AAFQE) was used during World War II for aircrew selection. Originally called the Aviation Cadet Qualifying Examination (ACQE) when initiated in 1942, it was used for preliminary selection of aircrew officers: pilots, bombardiers, and navigators. The name was changed when it was decided to use the test to select enlisted men for aircrew gunners. By 1947, the test had been administered to over a million men. It was primarily a power test with a three-hour time limit with a correction for guessing. Applicants could take the test as many times as they wished at 30-day intervals. It was also administered to civilians in high schools and colleges. One version of the test was described as consisting of 150 questions covering general vocabulary, reading comprehension, practical judgment, mathematics, contemporary affairs in aviation and the war, and mechanical comprehension (Davis, 1947). Aptitude testing concepts and psychometric techniques developed for aircrew personnel during World War II were carried forward and applied by the Air Force to classifying enlisted personnel in a broad array of job specialties after the war.

III. POST-WORLD WAR II TESTING DEVELOPMENTS

After the war, the Services including the Army Air Forces were allowed to develop their own selection tests for a few years. After being established as a separate Service branch in 1947, the Air Force continued to use forms of the AGCT, but they also wanted an enlistment screening test that would maximize the acceptance of airmen who would be able to meet classification test standards for enlisted specialties. This approach would enable the Air Force to choose those most qualified for their jobs. A new screening test called the Airman Qualifying Examination (AQE) was developed, and it had maximum correlation with the more valid tests in the Airman Classification Test Battery (ACTB). The test was designed to be essentially self-administering in a two-hour time period at the induction stations; most applicants could finish the test within 90 minutes (Research Bulletin No. 48-5, 1948). Although ready for use, the AQE was not implemented by the Air Force for screening for 10 years (see Table 1).

The delay in using the AQE was due to a major event in Department of Defense testing policy that occurred in 1948. A working group had been established to develop a single aptitude test for enlisted selection to be used by all Services. Some of the objectives were to minimize the importance of speeded tests, minimize the difficulty of verbal instruction, and include items in vocabulary, arithmetic reasoning, and spatial relations. The resulting test was the Armed Forces Qualification Test (AFQT) which was first used operationally in July 1950 (Eitelberg, Laurence, Waters, & Perelman, 1984). The Army was the executive agency for the development of the AFQT. Along with the Air Force, Navy, and Marines, they developed a 100 multiple-choice item test that covered vocabulary, arithmetic, spatial relations, and mechanical ability. It was the first test to be used for uniform screening of recruits across the Services (Gade, & Dudley, 2004). Supplementary tests were also used by the services. The Army Classification

Battery and the Army Qualification Battery were used by all Services but the Air Force from the early 1950's to the early 1970's.

While using the AFQT for screening, the Air Force developed Service-unique classification tests. Studies conducted at Lackland AFB,TX found that the AGCT did not measure many of the abilities that were required for success in Air Force jobs and that it overemphasized other measures that were less valid for Air Force success. A classification test for the Air Force called the Airman Classification Test Battery (ACTB) was developed and tested at Lackland AFB beginning in January 1947. The ACTB was later called the Airman Classification Battery (ACB). The combinations of tests from the classification battery were better at predicting Air Force specialty success than the Army General Classification Test was at predicting specialty success. It was further found that each Air Force enlisted specialty required its own pattern of aptitudes, but some of the patterns of aptitudes in specialties were similar and could be combined into homogeneous clusters or groups. Eight clusters of aptitudes were found to cover all of the airmen specialties. The composite scores for the clusters were converted to a stanine score for each group ranging from 1 to 9 with 5 representing the average score. Airman were assigned to a training specialty based on the minimum stanine score requirement for that specialty and what score they received. The term stanine score was later changed to the Aptitude Index as an equivalent and interchangeable term (Research Bulletin No. 48-4, 1948). The Air Force used the ACB series for classification from 1948 through 1958, as shown in Table 1.

IV. THE ARMED FORCES QUALIFICATION TEST (AFQT)

The AFQT was used as a joint service screening test from 1950 until 1973. Brokaw (1959) evaluated the AFQT for screening Air Force applicants. He found AFQT scores to be positively correlated with final technical training grades and Airmen Proficiency Tests and concluded that the test was effective for screening for success in Air Force specialties.

In 1958, the Services requested and received permission from Congress to add their own unique tests to supplement the AFQT. The Air Force used the Airman Qualifying Examination (see Table 1) which was in development prior to the initiation of the AFQT to help make selection and classification decisions (Maier, 1993). By 1972, the Services were allowed to use their own tests as long as they had conversion tables to the AFQT. In 1976, the Armed Services Vocational Aptitude Battery (ASVAB) was made the single Department of Defense test (Sellman, 1975).

AFQT Forms 1-8 were prepared in-house by the Army. The test originally included items to measure verbal skills, arithmetic reasoning, and spatial relations. A tool functions subtest was added in 1953 and dropped in 1973. In 1980, the spatial relations subtest was dropped from the aptitude tests, because too many illiterates who could qualify on the non-verbal items were subsequently failing the training courses (Maier, 1993). More emphasis was given to verbal and quantitative skills.

Aptitude tests are normed against an already existing score scale to ensure consistency in expected performance. The AFQT was normed against the AGCT. The AGCT was based on scores from the World War II population, so the qualification standards of the AFQT were tied to the characteristics of the World War II population. New forms of the AFQT and the Services' separate selection and classification batteries continued to be calibrated to the World War II population until the early 1960's (Maier, 1993). In 1960, the University of Pittsburgh and the American Institute for Research conducted a study to identify and define human talents. A comprehensive battery of tests was given to 400,000 high school students representing about 5% of the secondary schools in the United States. Called the TALENT battery, it was also given to a representative sample of Air Force enlisted men to calibrate the Air Force battery of aptitude tests: the Armed Forces Qualification Test (AFQT), the Airman Qualifying Examination (AQE), and the Air Force Officer Qualifying Test (AFOQT) against the project TALENT battery. By using multiple regression techniques, it was determined that there was close agreement between scores on project TALENT and on Air Force batteries (Dailey, Shaycroft, & Orr, 1962). Project TALENT provided an up-to-date sample to use for norming Air Force tests which was important since there were indications that the World War II mobilization population was outdated and that the abilities of youth had changed. For the Mechanical, General, and Electronics composites, the Air Force sample was found to be comparable to the general population of 18-year olds who scored in the upper quarter of the Airman Qualifying Examination as measured by Project TALENT. However, the Air Force sample scored lower on the Administrative composite than the sample from Project TALENT (Lecznar, & Tupes, 1963).

Prior to the All Volunteer Force (AVF), all applicants were given the AFQT for selection into the Service and then each Service gave their own classification test. This was called two-stage testing (Maier, 1993). After the end of the Vietnam War in 1973, use of the AFQT across the Services became optional. This presented a burden to the examining stations because they were administering different exams for different Services. With the advent of the AVF, it was necessary to recruit high quality recruits. In 1974, the DoD once again called for a single test for the Services which became the Armed Services Vocational Aptitude Battery (ASVAB) (Gade, & Dudley, 2004).

IV. APTITUDE TESTING FOR WOMEN IN THE MILITARY

In 1956, all Services used the AFQT for selection of men and women. Forms 1 and 2 of the AFQT became common tests for all Services. During development, they were standardized with a group of female enlistees and found to only have a slight discrimination against women. Forms 3 and 4 of the AFQT put more emphasis on mechanical training or experience and less on verbal skills. Most women were placed in clerical or administrative jobs, so forms 3 and 4 lost some validity for selection of women. The Sub-Panel on Coordination of Research in Personnel and Training, Panel on Personnel and Training, Committee on Human Resources named the Air Force as the executive agency for the development of a selection test battery for women. The battery was to have separate verbal and quantitative scores. The new test was called the Armed

Forces Women's Selection Test (AFWST) and yielded a verbal and quantitative score as well as a total score (McReynolds, 1956).

The AFWST was found to have face validity for women and enough difficulty to differentiate at the upper scoring levels. The tests were initially standardized against the AGCT and then also administered to WAF in basic training. Additionally, the WAF Enlistment Screening Test Forms 1 and 2 were developed for screening individuals who would then be sent to the Armed Forces Examining Stations (AFES) to take the AFWST. The screening test was shorter (40 items) and similar in content to the AFWST (McReynolds, 1956). The AFWST was used for testing women in the Air Force until 1974.

V. AIR FORCE CLASSIFICATION TESTS

From the late 1940's to 1973, the Air Force used a dual testing approach for selection and classification. One aptitude test was used for selection and another was used for classification. Beginning in 1950, all applicants were given the AFQT as an enlistment screening tool and Air Force classification tests to determine the best area of training and assignment for the recruit.

A fundamental postulate of classification is that each job requires a unique pattern of specific abilities for successful performance. In the Air Force, where there are hundreds of enlisted specialties, the ideal approach to classification would be to have a separate battery of tests predictive of performance in each job. Administratively, that is not feasible. Instead, early research was guided by the principle that there are jobs requiring similar patterns of aptitudes, and those jobs can be identified and combined into homogeneous clusters. Another research principle was that specific abilities predictive of individual performance in different job clusters can be identified, measured, and combined into composite scores. The aptitude composite scores would yield information for differentially predicting the utility of assigning each recruit to each job cluster. The difficulty of designing a classification test was highlighted by Brokaw (1960): "The selection of proper content for a classification battery for use in the assignment of Air Force enlisted personnel is a psychometric and philosophical problem of some magnitude. Psychometric complexity arises from the many criterion samples with biases resulting from differences in their original selection. Philosophical complexity arises from the attempt to devise a system of tests which is sufficiently stable to meet the needs of the Air Force personnel programs while being sufficiently flexible for use." Early research on enlisted classification was influenced by theories and statistical procedures developed by Brogden (1946, 1951, 1954), Horst (1954), Mollenkopf (1950), and Thorndike (1949).

Airmen qualified for entry into the military by taking the AFQT at recruiting stations across the country. In addition, the Air Force administered classification tests at basic training bases for assignment purposes. Between 1948 and 1975, the Air Force used different multiple aptitude batteries for the purpose of either classifying or selecting and classifying non-prior service enlistees. The first tests were a series of batteries

known as the Airman Classification Battery (ACB) followed by a series of batteries known as the Airman Qualifying Examination (AQE) (see Table 1). Brokaw and Burgess (1957) reported that the classification batteries were updated to protect from loss of security from repeated use, change test content that was obsolete, add new technology, and take advantage of advances in test theory. These classification tests are described in detail by Lecznar and Davydiuk (1960) and by Weeks, Mullins, and Vitola (1975). A summary of the ten classification tests used from 1947 to 1972 is provided in an appendix to this paper.

VI. THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB)

The Services recognized that the high schools were a rich source for military recruitment. Prior to 1962, no operational testing was done in the high schools to determine the potential aptitudes of students for military training. In 1962, a high school testing program was inaugurated by the Air Force Recruiting Service. It was felt that the test would be beneficial to both the Air Force and the schools. The test scores would provide valuable information about the characteristics of the high school enlistment pool and give high school counselors a tool to use to help the students make military career decisions. Other Services followed with their own aptitude testing batteries. The initial Air Force test was a form of the AQE and was calibrated against the Project TALENT norms. By 1968, the tests had been given to 400,000 students in 9,000 schools (Vitola, 1968).

In 1966, the Assistant Secretary of Defense for Manpower and Reserve Affairs requested a determination of the feasibility of using a common aptitude test battery that would "serve as an instrument for counseling high school students on vocational choices, could provide appropriate military service qualification data, and could be used in making job classification decisions about military enlistees" (Sellman, 1975) (Vitola, 1968).

A working group from all the Services developed the Armed Services Vocational Aptitude Battery (ASVAB) using the best parts of the various Services' tests. As a result of the DoD directive for a single test, the first ASVAB for student testing was introduced in 1968. With the advent of the All Volunteer Force, the student testing program became more important to the selection and classification of qualified military recruits. To strengthen the student testing program, the Armed Forces Vocational Testing Group (ASVTG) was initiated in 1972 to provide a single DoD manager for administering and scoring the high school test and provide better guidance for score interpretation. Because of criticism that the military use of the ASVAB scores for recruitment was not apparent, the Mosher Agreement was initiated in 1977. This agreement stated the principles and intent of the testing program and insured student privacy. Enhancements to the Student Testing Program have been made over the years to include counseling manuals, student workbooks, and other materials (Maier, 1993). Student participation is

voluntary and the students' names and test scores are provided to recruiters for each Service.

The AFQT was initiated for all Services in January, 1950 and was used until 1973 when the Services were allowed to select their own tests. At that time, the Air Force and Marines used a version of the ASVAB that was parallel to the battery used in the Student Testing Program, the Army used the Army Classification Battery, and the Navy used a Basic Test Battery. This presented a heavy demand on the examining stations that had to administer three different batteries each of which was over three hours in length and required separate testing facilities. These separate tests were used from 1973 through 1975. On January 1, 1976, all Services started using the Joint Service ASVAB. This reduced the burden on the examining stations and allowed applicants to take only one test before deciding on the branch of Service they would join. With the implementation of the ASVAB, all Services used it as one-stage testing for selection and classification. However, the Enlistment Screening Test and Computerized Adaptive Screening Tests were sometimes given to help determine who should go forward for ASVAB testing (Maier, 1993).

The first ASVAB used by the Air Force for selection and classification in 1973 was ASVAB-3. It replaced the AQE-J and the AFQT. ASVAB-1 had been accepted for use in the high school testing program and had been replaced in the high school program with ASVAB-2. Vitola and Alley (1968) developed ASVAB-3 using the test composites from ASVAB-1 for the Air Force Selection and Classification Program.

ASVAB-3 was composed of 9 subtests. For Air Force use, four composites were derived from the subtests and combined to form the indexes for Mechanical, Administrative, General, and Electronics (MAGE) composites. ASVAB-3 was similar to ASVAB-1. Composite reliabilities for ASVAB-1 ranged from .84 to .91 with a median of .89. Validities for ASVAB 1 ranged from .29 to .87 with a median of .68 (Weeks et al., 1975).

VII. THE EVOLUTION OF THE MAGE APTITUDE INDEXES

The development of the Airman Classification Batteries gave birth to the composite scores that are known as the Aptitude Indexes (AIs). The early aptitude batteries differed in number and types of abilities measured and the configuration of job clusters. As the tests were updated and refined, the number of AIs was reduced to four. The challenge was not only to predict success accurately in each job cluster, but also to accurately predict differences in success for each cluster. The AIs had to be both valid and differentially valid. If each of the separately developed AIs were equally valid for each job family, differential prediction, the core requirement for effective classification decisions, would be impossible. AIs with low intercorrelations allowed for differential prediction.

The first ACB, AC-1A, yielded eight AIs: Mechanical, Clerical, Equipment Operator, Radio Operator, Technical Specialty, Services, Craftsman, and Instructor. A

year later, the AC-1B replaced AC-1A and also yielded eight AIs. The Instructor AI was deleted, but an Electronics Technician AI was added and the battery had somewhat better differentiation. When AC-2A was introduced, the Air Force specialties had been combined into five major clusters and for the first time the MAGE composites (Mechanical, Administrative, General, and Electronics) appeared together. The fifth AI was Radio Operator and Electronics Technician became Electronics. When the AQE-D was introduced in 1958, the Aptitude Indexes had been reduced to the four composites that are still used today, MAGE (Weeks et al., 1975). The ASVAB also yields composite scores in the MAGE areas for the Air Force. Each Service has its own set of aptitude composite scores derived from the ASVAB using groups of ASVAB tests for the composites.

The configuration of the MAGE classification system was developed using a mix of expert judgments of job properties and similarities as well as the empirical relationships between the subtests and performance in Air Force training. Researchers relied on statistical methods including factor analysis and tests of correlation coefficients, which in the earliest studies were computed by hand.

From the mid-1950's through the 1980's, the test content continued to change, including the introduction of the Joint Service ASVAB in 1973 when several Air Force subtests were dropped and subtests developed by the other Services were adopted. Air Force subtests changed in content to reflect procedural updates and technology innovations. Over the years, advances were also made in analytical capabilities by the Air Force Human Resources Laboratory. One technique called hierarchical grouping (Ward, Treat, & Albert, 1984) provided a sophisticated new approach for job clustering. In a major study, the hierarchical grouping procedure was used to determine empirically-derived homogeneous clusters of Air Force entry-level jobs (Alley, Treat, & Black, 1988). Of special interest was whether the traditional four-group MAGE configuration would emerge from the empirical relationships.

Using technical school grades as the criterion and subtest scores on ASVAB Forms 8, 9, and 10 as predictors (Ree, Mathews, Mullin, & Massey, 1981), least squares regression equations were obtained for 211 Air Force specialties in the Alley et al., (1988) study. Then, using the hierarchical grouping method, predicted scores were generated for all recruits in the sample across all courses by applying the course-specific regression weights to each recruit's ASVAB subtest score. The resulting 211 technical school equations were then grouped on the basis of similarity of their predicted scores vectors, beginning with 211 separate equations and ending with a single consolidated equation. For comparison purposes, the four-group MAGE solution was also derived. This procedure resulted in four sets of specialties from which subtest least squares regressions were derived. The resulting equations were then compared with those obtained in the empirical solution to determine similarities in subtest weighting patterns (Alley et al., 1988).

The grouping solution for the last six stages of the clustering routine was examined. Three clusters corresponded approximately to the traditionally defined

Administrative, General, and Electronics job groups. A fourth cluster had a mix of mechanical maintenance and craftsman jobs. A fifth cluster was composed almost exclusively of job specialties with tactical and strategic aircraft engine maintenance. The final cluster was difficult to characterizes, because none of the career fields included was well predicted by subtests in the ASVAB (Alley et al., 1988).

The conclusions drawn were that the technical school regression equations revealed a pattern of job clusters and corresponding composites that, at the six-stage solution, yielded job groups/aptitude composites that clearly resembled the current MAGE system (Alley et al., 1988). However, the apparently enduring and robust nature of the MAGE distinctions did not obscure the fact that individual specialties had changed and reclassification from one to another of the MAGE categories was warranted. Two additional groups of jobs were defined. The one group, which was not predictable from the ASVAB, clearly indicted the need for additional research on the underlying requirements, not all of which appeared to be cognitive in nature. In the fifth group, jobs were complex and highly demanding and appeared to be those of a "generalist" instead of a "specialist." Further, some subtests in the ASVAB – particularly Numerical Operations, Coding Speed, and Mechanical Comprehension – had significant weights on few of the job clusters.

In 2002, the ASVAB was modified with the removal of the two speeded tests, Coding Speed (CS) and Numerical Operations (NO), and the addition of a spatial test, Assembling Objects. As a result, the Air Force reformulated those composites that formerly used CS and NO. The Air Force is currently evaluating the use of additional classification composites to supplement MAGE.

VIII. SUMMARY

Beginning with early development of the Army Alpha and Army Beta tests during World War I, the military has been instrumental in developing and refining the field of aptitude testing. The Army General Classification Test, developed during World War II as a test of general ability, became the capstone for a series of selection and classification tests that have been used by the Services since the 1940's. Additionally, the developments of the Army Air Force's Aviation Psychology Program contributed to the development of classification tests. Over the years, the Air Force has employed aptitude tests as required by the Congress and DoD and has, at times, supplemented the required tests with Air Force-unique tests to make personnel decisions.

For many years, the Air Force used a dual testing approach. From 1950 to 1973, the Services were required to use the AFQT as the enlisted selection instrument. The AFQT was used for selection and Air Force-developed tests were used for classification. Classification tests provided the ability to predict how an enlistee would perform in certain specialties. To build a test that would be unique to Air Force requirements, the Airman Classification Battery was implemented in 1948 with the first test, the AC-1A. The classification tests used combinations of subtests to derive composite scores that

could be used for prediction of success in various specialties. The AC-1A yielded eight composite scores, but as the tests evolved and became more refined, four composites for Mechanical, Administrative, General, and Electronics (MAGE) job specialties were defined as unique areas of job specialties. Even today under the single test system using the ASVAB, the Air Force derives scores for the MAGE composites to use in classification.

Personnel decisions are made when a person applies for enlistment, when the decision is made determining which job specialty the applicant qualifies for, and when the new recruit is assigned to a specialty. From the Army Alpha/Army Beta tests of World War I to the current ASVAB, the measurement of abilities with aptitude testing has contributed significantly to the quality of enlisted personnel and the capability of the Air Force to perform its mission.

REFERENCES

Alley, W.E., Treat, B.R. & Black, D.E. (1988). *Classification of Air Force jobs into aptitude clusters* (AFHRL-TR-88-14, AD-A206 610). Brooks AFB, TX: Air Force Human Resources Laboratory.

Brodgen, H.E. (1946). An approach to the problem of differential prediction. *Psychometrika, 11,* 139-154.

Brogden, H.E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement, 11,* 173-196.

Brogden, H.E. (1954). A simple proof of a personnel classification theorem. *Psychometrika, 19,* 205-208.

Brokaw, L.D. (1959). *Prediction of Air Force training and proficiency criteria from Armed Forces selection tests* (WADC-TN-59-194, AD-22). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Center.

Brokaw, L.D. (1959). *Prediction of Air Force training and proficiency criteria form Airman Classification Battery AC-2A* (WADN-TN-59-196). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Center.

Brokaw, L.D. (1960). *Suggested composition of airman classification instruments* (WADD-TN-60-214, AD-252 252). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division.

Brokaw, L.D. (1963). *Prediction of success in technical training from self-report information on educational achievement* (PRL-TDR-63-11, AD-414 888). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Brokaw, L.D. & Burgess, G.G. (1957). *Development of Airman Classification Battery AC-2A* (AFPTRC-TN-57-1, AD-131 422). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Dailey, J.T., Shaycroft, M.F. & Orr, D.B. (1962). *Calibration of Air Force selection tests Project TALENT norms* (PRL-TDR-62-6, AD-285 185). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division .

Davis, F.B.(Ed) (1947). *The AAF Qualifying Examination.* Report No. 6, Washington, D.C.: Office of the Air Surgeon, Headquarters, Army Air Forces.

Edwards, D.S & Hahn, C.P. (1962). *Development of Airman Qualifying Examination–62* (PRL-DR-62-7. AD-284 775), Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Eitelberg, M.J., Laurence, J.H., Waters, B.K., & Perelman, L.S. (1984). *Screening for service: Aptitude and education criteria for military entry.* Department of Defense: Office of Assistant Secretary of Defense (Manpower, Installations and Logistics).

Friedman, F.N. & Detter, H.M. (1954). *Factor analyses of the Airman Classification Battery AC-1A and selected Air Force and civilian tests from the 1949 normative sample* (AFPTRC-54-75). Lackland, TX: Air Force Personnel and Training Center.

Friedman, F.N. & Ivens, F.C. (1954). *Factor analysis of the Airman Classification Battery AC-1B, the USES General Aptitude Test Battery, experimental visualization and spatial tests, and psychomotor tests* (AFPTRC-TR-54-67). Lackland, TX: Air Force Personnel and Training Center.

Fruchter, D.A. (1963). *Development of Airman Classification Test-1963* (PRL-TDR-63-4, AD-404 039). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Gade, P.A. & Dudley, N.M. (2004). Sixty years of U.S. Army selection and classification test development. Brussels, Belgium: *Proceedings of the 46th Annual Conference of the Military Testing Association,* 791.

Headquarters Air Training Command (1948). *The development of an airman qualifying examination.* Barksdale AFB, LA: Research Bulletin No.48-5, Author.

Headquarters Air Training Command (1948). *The development of the Airman Classification Test Battery.* Barksdale AFB, LA: Research Bulletin No. 48-4, Author.

Horst, P.A. (1954). A simple proof of a personnel classification theorem. *Psychological Monographs, 68 (9),* Whole No. 380.

Judy, C.J. (1960). *Appraisal of educational requirements for airmen specialties* (WADD-TN-60-264, AD-252 253). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division.

Judy, C.J. (1969). Some highlights of military selection and classification testing in three wars. Governor's Island, New York: *Proceedings of the 11th Annual Conference of the Military Testing Association,* 392.

Lecznar, W. B. (1960). *Equivalence of scores from three airman classification devices* (WADD-TN-60-211, AD-245 431). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division.

Lecznar, W.B. (1961). Development of the Airman Classification Test - 1961 (ASD-TN-61-42, AD-261 502). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Systems Division.

Lecznar, W.B. (1962). *Some Aptitude data on Air Force enlisted accessions* (PRL-TDR-62-10, AD-289 874). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Lecznar, W.B. & Davydiuk, B.F (1960). *Airman classification test batteries: A summary* (WADD-TN-60-135, AD-240 831). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division.

Lecznar, W.B. & Tupes, E.C. (1963). *Comparison of Air Force aptitude indexes with corresponding TALENT test composites* (PRL-TDR-63-18, AD-420 555). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Madden, H.L & Lecznar, W. B. (1965). *Development and standardization of Air Force Qualifying Examination – 64* (PRL-TR-65-14, AD-622 807). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Madden, H.L., Valentine, L.D. & Tupes, E.C. (1966). *Comparison of the Airman Examination with the Differential Aptitude Test* (PRL-TR-66-7, AD-639 238). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Maier, M.H. (1993). *Military aptitude testing: The past fifty years* (DMDC Technical Report 93-007). Monterrey, CA: Defense Manpower Data Center.

Massey, I.H. & Creager, J.A. (1956). *Validation of the Airman Classification Battery: 1949-1953.* (AFPTRC-TN-56-129, AD-098 903). Lackland AFB, TX: Air Force Personnel and Training Research Center.

McReynolds, J. (1956). *Mental Qualification Tests for Women in the Armed Forces* (AFPRTC-TN-56-87). Lackland AFB, TX: Air Force Personnel and Training Research Center.

McReynolds, J. (1963). *Validity of Airman Qualifying Examination Form F, for technical training grades – 1961* (PRL-TDR-63-20, AD-426 756). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Mollenkopf, W.G. (1950). Predicted differences and differences between predictions. *Psychometrika, 15,* 409-417.

Ree, M.J, Mathews, J.J., Mullins, C.J., & Massey, R.H. (1981). *Calibration of Armed Services Vocational Aptitude Battery Forms 8, 9, and 10* (AFHRL-TR-81-49, AD-A-114 714). Brooks AFB, TX: Air Force Human Resources Laboratory.

Sellman, W.S. (1975). Use of common aptitude test for entry into all military services. Fort Benjamin Harrison, IN: *Proceedings of the 17th Annual Conference of the Military Testing Association,* 18.

Thompson, C.A. (1958). *Development of the Airman Qualifying Examination, Forms D and E* (WADC-TR-58-94(I), AD-151-045). Lackland AFB, TX: Wright Air Development Center.

Thorndike, R.L. (1949). *Personnel selection: Test and measurement techniques.* New York: Wiley.

Vitola, B.M. (1968a). An historical development of the military services high school testing program. San Antonio, TX: *Proceedings of the 10th Annual Conference of the Military Testing Association,* 30.

Vitola, B.M. (1968b). *Development and standardization of the Airman Classification Test – 1968* (AFHRL-TR-68-115, AD-687 090). Lackland AFB, TX: Air Force Human Resources Laboratory.

Vitola, B.M. & Alley, W.E. (1968). *Development and standardization of Air Force composites for the Armed Services Vocational Aptitude Battery* (AFHRL-TR-68-110, AD-688 222), Lackland AFB, TX: Air Force Human Resources Laboratory.

Vitola, B.M., Massey, I.H., & Wilbourn, J.M. (1971). *Development and standardization Of the Airman Qualifying Examination – Form J* (AFHRL-TR-71-28, AD-730 592). Lackland AFB, TX: Air Force Human Resources Laboratory.

Ward, J.H., Jr., Treat, B.R., & Albert, W.G. (1984). *General applications of hierarchical grouping using the HIER-GRP computer program* (AFHRL-TR-84-42, AD-A150 266). Brooks, AFB, TX: Air Force Human Resources Laboratory.

Weeks, J.L., Mullins, C.J., & Vitola, B.M (1975). *Airman Classification Batteries From 1948 – 1975: A review and evaluation* (AFHRL-TR-75-78, AD-026 470). Lackland AFB, TX: Air Force Human Resources Laboratory.

Wigdor, A.K. & Green, Jr., B.F., Eds (1991*). Performance Assessment for the Workplace Volume I.* Washington, D.C.: National Academy Press.

Related References

Humphries, L.G. (1962). *Stability of airman classification test scores* (PRL-TDR-62-3, AD-278 669). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Lecznar, W.B. (1959). *Preparation of the Airman Classification Test – 1960* (WADC-TN-59-197, AD-228 453). Lackland AFB, TX: Wright Air Development Center.

Lecznar, W.B. (1963). *Survey of tests used in airman classification* (PRL-TDR-63-5, AD-403 831). Lackland AFB, TX: Personnel Research Laboratory, Aerospace

Medical Division.

Lecznar, W. B. (1964). *Comparison of test items across forms* (PRL-TDR-64-3, AD-437 953). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

McReynolds, J. (1960). *Development of motivation keys for the Armed Forces Qualifications Test Forms 3 and 4* (AFPTRC-56-60). Lackland AFB, TX: Air Force Personnel and Training Center.

McReynolds, J. (1961). *Development of screening and enlistment tests for women* (ASD-TN-61-54, AD-266 865). Lackland AFB, TX: Aeronautical Systems Division.

Tupes, E.C. & Shaycroft, M.F. (1964). *Normative distributions of AQE aptitude indexes for high-school age boys* (PRL-TDR-64-7, AD-605 821). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Valentine, L.D. (1968). *Relationship between Airman Qualifying Examination and Armed Forces qualifying test norms* (AFHRL-TR-68-106, AD-678 528). Lackland AFB, TX: Air Force Human Resources Laboratory.

Appendix
Air Force Classification Tests (1948-1973)

Airman Classification Battery AC-1A

The first Air Force Classification Battery, AC-1A, was initiated operationally in November, 1948. It consisted of 12 aptitude tests and a Biographical Inventory and took five hours and twenty minutes to administer. The test yielded eight composite scores or aptitude indexes (AIs). The AIs were Mechanical (M), Clerical (Cl), Equipment Operator (EO), Radio Operator (RO), Technician Specialty (TS), Services (S), Craftsman (Cr), and Instructor and (I). Some of the tests were based on tests developed by the Aviation Psychology Program and some were developed to measure performance required by specific job clusters. Reliabilities for the Battery were high enough to use it for classification purposes. The test/retest reliabilities ranged from .89 to .96 with a median of .92. Validities showed a positive relationship between the test scores and technical training school success. The validities of the indexes ranged from .32 to .77 with a median of .61. To be most effective the aptitude indexes must be valid differentially. Without differentiation there is no need for separate indexes. The intercorrelations for the aptitude indexes were not optimal, ranging from .50 to as high as .91 with a median of .81. There appeared to be too much correlation among the AIs for the test to have strong differential prediction. Ultimately, about 120,000 basic airmen were tested for classification to Air Force jobs (Weeks, Mullins, & Vitola, 1975).

In 1949, tests were administered to basic airmen to compare the Airman Classification Battery, AC-1A, with similar civilian tests and to compare the norms of the civilian tests with those of the airman population. This became known as the 1949 Normative Survey Battery. The civilian tests were the Terman-McNemar Test of Mental Ability, Form C; the Modified Alpha Examination, Form 9; the Cattell Culture-Free Tests; and Parts II and VI of the Guilford-Zimmerman Atitude Survey, Form A. Of the seven factors found to be common to all of the batteries, the Airman Classification Battery best measured verbal comprehension, mechanical experience, numerical facility, perceptual speed, and academic information. General reasoning and visualization were best defined by the Guilford-Zimmerman Aptitude Survey, but were also strong for the Airman Classification Battery (Friedman, & Detter, 1954).

After the Air Force started to administer the Airman Classification Battery (ACB) to new recruits, it was decided to have classification scores on as many personnel as possible. A short form of the Airman Classification Battery dated March, 1949 was used in decentralized test administration and was called the Airman Classification Test Battery – Permanent Party -1 (ACTB-PP-1). It was an exact duplicate form of the Airman Qualifying Examination (AQE-Form A). Form A had been printed in August 1948 and meant to be used as a screening test to accompany the ACB and as a substitute for the Army General Classification Battery for Air Force personnel. The items were designed to yield maximum accuracy of measurement at the lower levels of ability and the test

could be self-administered and easily scored. Form A was never used as a screening test (Lecznar, & Davydiuk, 1960).

Airman Classification Battery AC-1B

In December 1949, AC-1B replaced AC-1A with some minor changes and to respond to a need for an Electronics AI. The Instructor AI was dropped.. Research had indicated that separating the Electronics area from the Mechanical would result in better prediction, so an Electronics Technician AI was added. AC-1B consisted of 13 subtests with the addition of Pattern Comprehension subtest and a Biographical Inventory. Reliabilities on test/retest ranged from .68 to .93 with a median of .90 which was slightly lower that the AC-1A. Validity coefficients ranged from .34 to .77 with a median of .60 which was very similar to the validities for first test. The index intercorrelations ranged from -.06 to .85 with a median of .78. AC-B1 did have slightly better differentiation between composites than AC-A1.

AC-1B was revised in January 1953. At that time, the Radio Operator field was experiencing a high rate of attrition mainly due to difficulties in learning the International Morse Code. The Radio Operator AI was revised to include measures of code learning and numerical and verbal tests. It was revised again in April 1955 when the Services AI was dropped because it was showing very low predictive ability. (Weeks et al., 1975)

Friedman and Ivens (1954) compared the AC-1B to the United States Employment Service General Aptitude Test Battery (USES GATB). The study identified the common factor loadings of the AC-1B and the GATB. The tests were given to 190 basic airmen at Lackland AFB, TX during April and May of 1950. The AC-B1 was found to have significant factor loadings on Mechanical Experience, Numerical Facility, Verbal Comprehension, and Perceptual Speed factors. The AC-B1 had no measure of Psychomotor Speed and Psychomotor Coordination and Precision, but the GATB measured these abilities.

Airman Classification Battery AC-2A

AC-2A became operational in January 1956 with some major changes. It was decided that fewer aptitude indexes and a different grouping of Air Force jobs was needed. The new test had 15 subtests and yielded five aptitude indexes. The indexes were Mechanical (M), Administrative (A), Radio Operator (RO), General (G), and Electronics (E). The new index of General replaced the Technical Specialty of AC-B1. The goal was to produce an instrument with maximum differential validity. The Electronics Technician AI became the Electronics AI. This battery required about five hours and thirty minutes of testing time. The AC-2A, like its predecessors, was also standardized against the World War II population. Brokaw and Burgess (1957) reported that AC-2A was the first battery to group specialties into aptitude clusters using mathematical analyses instead of the judgments of job analysts. Scoring for this battery was also changed from stanines to percentiles.

This version was validated against technical school final grade and against job proficiency as measured by the Airman Proficiency Test. Validities for final school grade ranged from .11 to .80 with a median score of .57. Validities for job performance measured by the Airman Proficiency Test ranged from .19 to .75 with a median score of .58. Test/retest reliabilities ranged from .87 to .93 with a median of .89. Intercorrelations for differential validity show a range of -.02 to .81 with a median of .57. The discrimination between tests in AC-2A increased the differential validity over other versions. This battery provided improved composite predictors over the first two batteries (Weeks, Mullins, & Vitola, 1975).

In a study of the early classification tests, Massey and Creager (1956) said that there was improvement in manpower efficiency as a result of using the battery for classification. The battery had excellent coverage of verbal, numerical, and mechanical knowledge, but needed to do a better job of identifying spatial and reasoning abilities. The indexes showed high inter-correlations and therefore did not have much differential validity.

Airman Qualifying Examination, Form D – AQE-D

The AQE had been used in the Air Force since late 1948 as a short version of the classification battery. Forms D and E were essentially equivalent batteries and produced four aptitude indexes: Mechanical (M), Administrative (A), General (G), and Electronics (E). Form E was developed as an alternate form to Form D but was not used operationally as Form E (Lecznar & Davydiuk, 1960). The AQE was used in cases where the Airman Classification Battery was unavailable or inappropriate; for retest when there was something wrong with the airman classification test data collected on an airman; to prevent reenlistment when potential for retraining was minimal; to determine eligibility of prior service personnel for assignment to technical schools; and, beginning in April 1958, to select airmen at Armed Forces Examining Stations that fit the Air Force requirements. AQE-D was designed to produce aptitude indexes that were interchangeable with the current classification battery, but they did not produce all the indexes that the Airman Classification Battery produced. There was not a Radio Operator AI in the AQE (Thompson, 1958). Lecznar (1960) compared scores on an experimental sample who took the AC-2A, AQE-D and AQE-E and found the scores to be comparable across tests.

Madden, Valentine, and Tupes, (1966) studied the relationship between AQE scores and the Differential Aptitude Test (DAT) scores. The DAT is a commercial test used for vocational counseling. They found a positive relationship in predicting some of the AQE aptitude indexes from some of the weighted DAT subtests. Data indicate that the two test batteries measure essentially the same areas except for clerical speed which is not covered in the AQE.

The Air Force implemented a selective recruiting program at the recruiting stations in 1958 to ensure that the best applicants were selected from the applicant pool. They began using the AQE in April 1958 for selection and classification of non-prior

service applicants for selective enlistment. Prior to the selective recruiting program, Air Force applicants had to qualify on the AFQT, but the Air Force also needed an instrument that could be used in the field for both selection and classification. The administration time of the AC-2A was about a day, which was too long for field administration. They required a test of 4 hours or less. Additionally the scoring had been by machines, but the field locations required hand scoring.

The AQE-A had been designed as a short version of AC-1A to be used for screening. It was never used operationally for screening; but in 1949, it was issued as the Airman Classification Test Battery-Permanent Party Personnel – 1 (ACTB-PP-1). It was used to obtain Aptitude Indexes on Permanent Party personnel who had entered service before implementation of the AC-1A. Later it was redesignated AQE-A and used for retesting needs. Short forms AQE-B and AQE-C were used from March 1953 to September 1956.

AQE-D was implemented in April 1958 as a test that was comparable to the AC-2A . It was a two 2 hour and 15 minute, hand scored battery of 11 aptitude tests. Composite correlations between the AQE-D and the AC-2A ranged from .76 to .83 with a median of .81 indicating that they are measuring similar functions. The instrument was validated for only one job specialty in three of the four composites. These validities were significant but the overall validity data are insufficient for interpretation. Differentiation between the composites was about the same for AQE-D as for AC-2A (Weeks et al., 1975).

Air Qualifying Examination, Form F – AQE-F

In May 1959, it was decided to continue the use of the selective enlistment program, so the shorter AQE format was still needed. Introduced in November 1960 with 11 aptitude tests, the AQE-F was actually the already developed AQE-E with a few minor changes. The mechanical operations test was dropped from the Mechanical Index and the Tool Functions test was dropped from the Administrative Index. Also the Hidden Figures replaced Figure Recognition.

Test/retest reliabilities ranged from .81 to .88 with a median score of .83. Validities predicting technical school grades ranged from .28 to .90 with a median of .63. Since the AQE was being used for both selection and classification, some of the differential validity for classification was lost to achieve maximum total test validity (Weeks et al., 1975).

For all personnel classification programs that were not part of the selective enlistment program, the ACT – 61 was developed. It was a single-form, four-hour test with 10 subtests that yielded four aptitude scores: Mechanical, Administrative, General, and Electronics. It was comparable to AC-2A in internal consistency and difficulty. (Lecznar, 1961) Documentation of later Airman Classification Tests used for other than the selective enlistment program can be found in Fruchter (1963) and Vitola (1968).

McReynolds (1963) looked at the validity of the AQE-F for predicting technical training grades. Validities were determined for the four composites in 49 airman training courses. It was found to be an effective instrument for assigning airmen to technical training. The highest validity was with the electronics index and the lowest was with administrative. The results of this study indicated that the AQE-F was a good instrument for the assignment of enlistees to technical training.

Airman Qualifying Examination – 1962 – AQE-62

AQE-62 replaced AQE-F in October of 1962. The major change in the new test was the arrangement of the items in a spiral omnibus format. The AIs remained the same with the exception that the Airman Arithmetic subtest replaced the Clerical Matching subtest and the Numerical Operations subtest in the Administrative AI. The ten-item test battery required two hours for administration. As in previous tests, the standardization of the test was based on the World War II population.

The test/retest reliabilities ranged from .78 to .83 with a median of .80. Validities were inferred from the relationship of AQE-62 to AQE-F. The validity coefficients obtained were Mechanical .75, Administrative .76, General .81, and Electronics .81. Reliability coefficients were acceptable but there was a noticeable drop in reliability for the Administrative AI from .88 in AQE-F to .77 in AQE-62. (Weeks, Mullins, & Vitola, 1975). Edwards and Hahn (1962) reported that the AQE-62 closely paralleled the AQE-F.

Airman Qualifying Examination – 1964 (AQE-64)

The AQE-64 was introduced in November 1964 with several modifications. It was comprised of ten subtests in two booklets. One booklet was for the power tests and one was for the speeded tests. Arithmetic Computation replaced Airman Arithmetic, because test takers were taking too much time completing the Airmen Arithmetic questions that had been integrated with other test variables and they were not getting to the end of the battery. As recommended in research by Judy (1960) and Brokaw (1973), points for completion of academic courses were added to the AIs. The same four aptitude composites were derived from the subtests, but bonus points based on completion of five academic courses (algebra, geometry, trigonometry, physics, and chemistry) were added to get the composite scores. In July 1974, credit in the composites for high school courses was discontinued because only minor differences were found in composites that included credit for courses and composites that did not include credit (Vitola, & Alley, 1968). Pacing directions were also added to the administration instructions for AQE-64. Items covering Hidden Figures, Technical Data Interpretation, and Pattern Comprehension were spiraled into one test.

The Project TALENT sample, as discussed earlier in this paper, was first used for Air Force standardization purposes in this classification battery. The norms for 12th grade males in the TALENT sample were used as the normative reference base for the development of AQE-64.

Although reliability coefficients were not available, they were estimated to range from .80 to .90 based on the similarity of AQE-64 with AQE-62. Validities based on technical school course grades for airmen in 57 separate technical school courses ranged from .34 to .87 with a median of .64. Reliability coefficients for the AIs between AQE-62 and AQE-64 were .76 for Mechanical, .80 for Administrative, .82 for General, and .78 for Electronics (Weeks et al., 1975). In-depth information on the development of AQE-64 can be found in Madden & Lecznar, (1965).

Airman Qualifying Examination – 1966 (AQE-66)

The AQE-66 replaced the AQE-64 in September 1966. It was very similar to the AQE-64 with 10 subtests presented in two parts. Part I, the speeded test Arithmetic Computation, was moved to the first of the battery for ease of administration. Part II contained the power tests.

The test was standardized to the Project TALENT sample. Test/retest reliability for the composites ranged from .84 to .88 with a median of .87. The validities predicting technical school course grades ranged from .18 to .90 with a median of .68 (Weeks et al., 1975).

Airman Qualifying Examination Form J (AQE-J)

The AQE-J with ten subtests replaced the AQE-66 in July 1971 to prevent test obsolescence and compromise. The spiral omnibus format was dropped in favor of distinct subtest format. Test variables and composites mirrored AQE-66. Reliabilities for the composites ranged from .88 to .94 with a median of .91. Validities were inferred from the relationship of Form J to AQE-66. Correlation coefficients for the AIs of the two batteries were .82 for Mechanical, .69 for Administrative, .83 for General, and .84 for Electronics (Weeks et al., 1975). An in-depth description of the development of Form J can be found in Vitola and Wilbourn (1971).